An existence proof for the singular value decomposition
and its use for concept search, as given in the 2nd lecture of the seminar
Advanced Topics in Information Retrieval, WS 04/05

Holger Bast

October 27, 2004

Definition: Let $B$ be an $n \times n$ matrix, and let $x$ be a nonzero $n$-vector such that $B \cdot x = \lambda \cdot x$, for some real number $\lambda$. Then $x$ is called an eigenvector of $B$, and $\lambda$ is called the corresponding eigenvalue.

In the lecture, we assumed the following well-known theorem, which we did not prove (the first exercise on the second sheet goes a long way towards explaining though why this holds).

Theorem: For any symmetric $n \times n$ matrix $B$, that is $B_{ij} = B_{ji}$ for all $i$ and $j$, there exist $n$ eigenvectors $v_1, \ldots, v_n$, which are pairwise orthogonal, that is, $v_i^T v_j = 0$, for $i \neq j$. Moreover, $B$ can be written as $VDV^T$, where $V = [v_1 \ldots v_n]$ is the $n \times n$ matrix with the normalized(!) eigenvectors written column by column, $D$ is an $n \times n$ matrix containing the eigenvalues corresponding to $v_1, \ldots, v_n$ on its diagonal, and $V^T$ is just the transpose of $V$, that is, the normalized eigenvectors written row by row.

Remark: In the following we use that for any $n \times n$ matrix $M$ with normalized orthogonal columns, both $M^T M$ and $M M^T$ yield the $n \times n$ identity matrix — such a matrix is called orthogonal. $M^T M = I$ is clear, because that just says that the columns are normalized and pairwise orthogonal. Multiplying this by $M$ from the left gives $M M^T M = M$, and multiplying this by the inverse of $M$ from the right gives $M M^T = I$, which says that also the rows of $M$ are normalized and pairwise orthogonal. Note that $M$ has an inverse because by assumption its $n$ columns are orthogonal, and in particular hence linearly independent.

From this we derived the so-called singular value decomposition (SVD) of an arbitrary $m \times n$ matrix $A$ in a number of steps as follows:

1. We first observed that for an arbitrary $m \times n$ matrix $A$, the $m \times m$ matrix $AA^T$ as well as the $n \times n$ matrix $A^T A$ are symmetric.

2. By the theorem above $A^T A$ then has $n$ pairwise orthogonal eigenvectors. Let $v_1, \ldots, v_r$ be those normalized such eigenvectors pertaining to the nonzero eigenvalues $\lambda_1, \ldots, \lambda_r$ of $A^T A$. We observed that because by definition eigenvectors pertaining to an eigenvalue zero are mapped to zero, $r$ is just the rank of $A^T A$, which is also the rank of $A$.\footnote{In the lecture we actually started from the other matrix $AA^T$ but that turned out a bit unfortunate with the notation, so here I do it the other way round again, as I first did it myself, I only changed that one hour before the lecture because I actually thought that would help with the notation, sorry for that confusion . . .}
3. We then observed that $AA^T Av_i = A \lambda_i v_i = \lambda_i Av_i$, that is, that $v_i$ is an eigenvector of $A^T A$ implies that $Av_i$ is an eigenvector of $AA^T$ with the same eigenvalue $\lambda_i$. We computed the norm of $Av_i$ via

$$(Av_i)^T (Av_i) = v_i^T A^T Av_i = v_i^T \lambda_i v_i = \lambda_i v_i^T v_i.$$  

Since for any vector $x$, the squared norm $||x||^2 := \sum_i x_i^2$ is just the scalar product with itself: $x^T x$, we have that the norm of $Av_i$ is the square root of $\lambda_i$, which we denoted by $\sigma_i$.

4. Now define $u_i = Av_i/\sigma_i$; by the above $u_1, \ldots, u_r$ are then unit vectors just like the $v_1, \ldots, v_r$. Then, by the same argument as used for the previous displayed formula,

$$u_i^T Av_j = (Av_i/\sigma_i)^T Av_j = v_i^T A^T Av_j/\sigma_i = v_i^T v_j \cdot \lambda_j/\sigma_i,$$

that is, $u_i^T Av_j$ is zero for $i \neq j$ (because $v_1, \ldots, v_r$ are pairwise orthogonal), and $\sigma_i$ for $i = j$ (because the $v_1, \ldots, v_r$ are unit vectors and $\lambda_i = \sigma_i^2$).

5. Writing all these equations, for $i, j = 1, \ldots, r$, in matrix form gives

$$\begin{bmatrix} u_1^T \\ \vdots \\ u_r^T \end{bmatrix} \cdot A \cdot [v_1 \cdots v_r] = \text{diag}(\sigma_1, \ldots, \sigma_r),$$

where the right hand side denotes an $r \times r$ matrix with $\sigma_1, \ldots, \sigma_r$ on the diagonal and all zeros elsewhere.

6. The $u_i$ are vectors from an $m$-dimensional space, and we only have $r$ of them, so we can pick unit vectors $u_{r+1}, \ldots, u_m$ that are pairwise orthogonal, and orthogonal to $u_1, \ldots, u_r$. Similarly we can find $v_{r+1}, \ldots, v_n$ such that $v_1, \ldots, v_n$ are pairwise orthogonal unit vectors. We then still have that $u_i^T Av_j$ is $\sigma_i$ when $i = j \leq r$ and by a calculation analogous to the above it is zero for all other cases. In matrix form this gives us

$$\begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix} \cdot A \cdot [v_1 \cdots v_n] = \Sigma,$$

where $\Sigma$ is now an $m \times n$ matrix with its first $r$ diagonal entries being the $\sigma_1, \ldots, \sigma_r$ and zeroes everywhere else.

7. Defining $U = [u_1 \cdots u_m]$ and $V = [v_1 \cdots v_n]$, that is, writing the $u_1, \ldots, u_m$ resp. the $v_1, \ldots, v_n$ column by column, we now have $U^T AV = \Sigma$. By the remark following the eigenvector decomposition theorem stated in the beginning, $UU^T = U^T U = I$ and also $VV^T = V^T V = I$, and we thus get

$$A = U \Sigma V^T.$$

This is the famous singular value decomposition, and you have just seen a complete proof of its existence for an arbitrary $m \times n$ matrix $A$. 

2
8. Since $\Sigma$ has so many zeroes, all columns of $U$ and $V$ beyond the $r$th actually play no role in the matrix multiplication in $U\Sigma V^T$ (but we needed these columns to get from $U^TAV = \Sigma$ to $A = UAV^T$). Taking $U_r = [u_1 \cdots u_r]$, $V_r = [v_1 \cdots v_r]$, and $\Sigma_r = \text{diag}(\sigma_1, \ldots, \sigma_r)$, we hence also have the decomposition

$$A = U_r \Sigma_r V_r^T.$$ 

This is sometimes called the “economy” version of the SVD because, in contrast to the full version, there is no redundancy in it.

Remark: The $\sigma_1, \ldots, \sigma_r$, which are just the square roots of the eigenvalues of $AA^T$ resp. $A^TA$, are called the singular values of $A$. The columns of $U$, which are just the eigenvectors of $AA^T$, are called left singular vectors of $A$, and the columns of $V$, which are just the eigenvectors of $A^TA$, are called the right singular vectors of $A$.

The singular value decomposition (SVD) gives us one way to obtain a decomposition of the type which we found so useful in the first lecture and exercise. Namely, taking $U_k = [u_1 \cdots u_k]$, $V_k = [v_1 \cdots v_k]$, and $\Sigma_k = \text{diag}(\sigma_1, \ldots, \sigma_k)$, gives us the decomposition $U_k \Sigma_k V_k^T$ or $U_k \Sigma_k V_k^T$; in both cases an $m \times k$ matrix times a $k \times n$ matrix, as desired.

The theorem by Eckart and Young (from 1936) gives an indication of why this decomposition might be useful. The theorem says that no other matrix of rank at most $k$ — and each product of an $m \times k$ matrix with a $k \times n$ matrix has rank at most $k$ — is closer to $A$ in $L_2$ norm then $U_k \Sigma_k V_k^T$.

Proving this theorem is the second exercise of the second exercise sheet. It is a sequence of basic algebraic manipulations quite similar to the ones used in the derivations above. In doing this proof you will get acquainted with the eigenworld, no lecture can do that for you. It is a fascinating world (at least for those who love elegant mathematics), and, as should turn out more and more in the course of this seminar, also a very useful one to have one’s hands on . . .